Collège Sciences et Technologies
UF Mathématiques et Interactions

Le problème de p-centre

Ahina Durrieu
Maïna Boivent

Master MAS parcours ROAD

2024 – 2025

Projet d'optimisation

Table des matières

Engagement de non plagiat

Nous, Ahina Durrieu, Maïna Boivent, déclarons être pleinement conscients que le plagiat de documents ou d'une partie d'un document publiés sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, nous nous engageons à citer toutes les sources que nous avons utilisées pour produire et écrire ce rapport.

Fait à Talence le 21 janvier 2025

Signatures
Ahina Durrieu
Maïna Boivent

1 Introduction

Pour chaque territoire, il est essentiel, à partir d'un ensemble d'emplacements disponibles, de déterminer l'emplacement optimal des installations fournissant un service. L'objectif est de garantir un niveau de qualité de service tout en respectant une contrainte budgétaire.
Ce type de problématique est crucial pour le positionnement de services géographiques tels que les hôpitaux, les ambulances, les casernes de pompiers, les postes de police, les crèches, les écoles ou encore les bibliothèques.

Ces problématiques, qui visent à identifier les meilleurs emplacements pour des installations (ou ressources) ainsi qu'à allouer les clients (ou la demande) à ces installations, relèvent de la catégorie des problèmes de localisation-allocation. Il existe aujourd'hui une grande diversité de problèmes et de variantes dans cette catégorie. Ces questions sont fondamentales dans des domaines tels que le géomarketing, la logistique, l'aménagement du territoire et les transports.

Le problème de p-centre, que nous allons aborder dans ce projet, est un problème de type "pull". Il s'agit de déterminer l'emplacement des installations de manière à minimiser la distance maximale entre un client et son installation.
Concrètement, étant donné un ensemble d'emplacements potentiels pour accueillir les installations et un ensemble de clients, le problème de p-centre consiste à réduire au minimum la plus grande distance qui sépare un client de son installation.
Différentes versions du problème de p-centre seront modélisées et par la suite, des expérimentations seront menées pour comparer l'efficacité et la qualité des différentes formulations.

Le rapport est organisé de la manière suivante. La section 2 présente la formalisation du problème de p-centre. La section 3 présente trois formulations du problème basées sur la programmation linéaire, en variables mixtes ou en nombres entiers, d'abord sans prendre en compte les demandes et les capacités (section 3.1), puis en les prenant en compte (section 3.2). Nous introduisons plusieurs algorithmes de résolution du problème dans la section 4. Nous présentons les résultats des tests effectués sur les différents jeux de données fournis ainsi que leur analyse dans la section 5. Enfin, la section 6 est dédiée à une conclusion sur le projet.

2 Formalisation du problème

Le problème de p-centre (pCP) se définit formellement comme suit. Soit $G = (\mathcal{F}, \mathcal{C}, \mathcal{E})$ un graphe biparti. Soit m un entier positif représentant le nombre de nœuds candidats pour les installations, $\mathcal{F} = \{1, \ldots, m\}$ l'ensemble de ces nœuds. Soit n un entier positif définissant le nombre de nœuds représentant les clients, $\mathcal{C} = \{1, \ldots, n\}$ l'ensemble de ces nœuds. L'ensemble $\mathcal{E} = \{(i, j) : i \in \mathcal{F}, j \in \mathcal{C}\}$ est l'ensemble des arêtes. Chaque arête $(i, j) \in \mathcal{E}$ représente une affectation possible d'un client j à une installation i, associée à une distance réelle positive d_{ij}.

L'objectif du pCP est d'ouvrir au plus p installations, et d'affecter chaque client à exactement une installation, tout en minimisant la distance maximale entre un client et son installation.

Le problème de p-centre capacitaire (CpCP) est une variante qui reprend les données du problème de pCP, et qui ajoute les données suivantes : chaque installation $i \in \mathcal{F}$ a une capacité Q_i, qui peut être utilisée pour répondre aux demandes q_j des clients $j \in \mathcal{C}$.
L'objectif du CpCP est le même que celui du pCP, en ajoutant la nécessité de s'assurer que la demande totale des clients affectés à chaque installation ne dépasse pas la capacité de cette dernière.

Sauf indication contraire, nous supposons que tous les paramètres d'entrée (à l'exception des distances) sont des entiers.

Dans la suite de ce travail, une solution sera représentée par une liste d'installations ouvertes ainsi qu'une liste d'affectations des clients aux installations.

3 Modélisation du problème

Nous présentons ici la modélisation du problème de p-centre et de p-centre capacitaire, en utilisant la programmation linéaire, en variables mixtes et en nombres entiers. Dans tous les modèles suivants, les variables binaires x_i prennent la valeur 1 si l'installation $i \in \mathcal{F}$ est ouverte, et 0 sinon.

3.1 Formulations sans prendre en compte les demandes et capacités (pCP)

Dans un premier temps, nous ne prenons pas en compte les demandes des clients et les capacités des installations. Cela repose sur l'hypothèse que les installations ouvertes ont une capacité infinie.

3.1.1 Version 1 : Formulation classique ($p\text{CP}_1$)

Dans cette première version, nous allons modéliser le problème en utilisant la programmation linéaire en variables mixtes. Nous introduisons les variables binaires y_{ij} valant 1 si le client $j \in \mathcal{C}$ est affecté à l'installation $i \in \mathcal{F}$ et 0 sinon, ainsi que la variable réelle D prenant la valeur de la distance maximale entre un client et son installation.

$$
\begin{aligned}
[P1] \quad & \min D && (1)\\
& \text{s.c.} \quad \sum_{i \in \mathcal{F}} x_i \leqslant p && (2)\\
& \sum_{\substack{i \in \mathcal{F},\\ (i,j) \in \mathcal{E}}} y_{ij} = 1 \qquad \forall j \in \mathcal{C} && (3)\\
& y_{ij} \leqslant x_i \qquad \forall i \in \mathcal{F},\ \forall j \in \mathcal{C} && (4)\\
& D \geqslant d_{ij} y_{ij} \qquad \forall (i,j) \in \mathcal{E} && (5)\\
& x_i \in \{0,1\} \qquad \forall i \in \mathcal{F} && (6)\\
& y_{ij} \in \{0,1\} \qquad \forall j \in \mathcal{C} && (7)\\
& D \in \mathbb{R}^+ && (8)
\end{aligned}
$$

L'objectif (1) consiste à minimiser la plus grande distance entre un client et son installation.
La contrainte (2) impose d'ouvrir au plus p installations.
Les contraintes (3) imposent que chaque client soit affecté à exactement une installation.
Les contraintes (4) assurent que l'on n'affecte pas un client à une installation pas ouverte. Les contraintes (5) assurent que D est la valeur de la distance maximale entre un client et son installation : si D est supérieur à chaque distance, alors il vaut le maximum des distances.
Les contraintes (6), (7) et (8) définissent les domaines des variables de décision.

3.1.2 Version 2 : Première formulation basée sur le rayon de couverture ($p\text{CP}_2$)

Dans cette deuxième version, nous allons utiliser le rayon de couverture. Le rayon de couverture optimal se définit par la plus petite distance D^* permettant d'affecter chaque client à une installation avec une distance inférieure ou égale à D^*.
On écrit alors $D^0 < D^1 < ... < D^K$ les différentes valeurs de d_{ij} pour tout $(i,j) \in \mathcal{E}$, et ainsi les valeurs possibles du rayon de couverture optimal. On note $\mathcal{K}$ l'ensemble $\{1, \ldots, K\}$.
Nous introduisons la variable binaire z_k pour chaque $k \in \mathcal{K}$, qui vaut 1 si le rayon de couverture est supérieur ou égal à D^k et 0 sinon, et nous n'utilisons plus les variables y_{ij}.

On précise que la variable z_0 n'est pas définie, car elle vaut toujours 1, le rayon de couverture optimal étant obligatoirement supérieur ou égal à la plus petite distance.

$$[P2] \quad \min D^0 + \sum_{k \in \mathcal{K}} (D^k - D^{k-1}) z_k \tag{9}$$

$$\text{s.c. } (2), (6)$$

$$z_k \geqslant 1 - \sum_{\substack{i \in \mathcal{F}, \\ d_{ij} < D^k}} x_i \qquad \forall j \in \mathcal{C}, \ \forall k \in \mathcal{K} \tag{10}$$

$$z_k \in \{0, 1\} \qquad \forall k \in \mathcal{K} \tag{11}$$

L'objectif (9) consiste à trouver le rayon de couverture optimal. Tant que z_k vaut 1, on additionne la différence entre D^k et D^{k-1}, de sorte qu'à l'indice k, la somme vaut exactement D^k (si z_k vaut 1). Ainsi, si z vaut 0 à partir de l'indice $k+1$, alors le rayon de couverture optimal est D^k, et c'est bien ce que vaut la somme.
Les contraintes (10) imposent que z_k vaut 1 si le nombre d'installations ouvertes dans le rayon D^{k-1} pour un client j est de 0. Sinon, si il y a au moins une installation dans ce rayon pour chaque client, alors z est libre et vaudra donc 0 car l'objectif est en minimisation. Ainsi, z vaut bien 1 pour tous les rayons inférieurs ou égaux au rayon optimal, et 0 pour les autres.
Les contraintes (11) définissent les domaines des variables de décision.

3.1.3 Version 3 : Deuxième formulation basée sur le rayon de couverture (pCP$_3$)

Dans cette troisième version, nous utilisons encore le rayon de couverture. On introduit cette fois la variable binaire u_k pour chaque $k \in \mathcal{K} \cup \{0\}$, qui vaut 1 si D^k est le rayon de couverture optimal, et 0 sinon. Nous n'utilisons plus les variables z_k.

$$[P3] \quad \min \sum_{k \in \mathcal{K} \cup \{0\}} D^k u_k \tag{12}$$

$$\text{s.c. } (2), (6)$$

$$u_k \leqslant \sum_{\substack{i \in \mathcal{F}, \\ d_{ij} \leqslant D^k}} x_i \qquad \forall j \in \mathcal{C}, \ \forall k \in \mathcal{K} \cup \{0\} \tag{13}$$

$$\sum_{k \in \mathcal{K} \cup \{0\}} u_k = 1 \tag{14}$$

$$u_k \in \{0, 1\} \qquad \forall k \in \mathcal{K} \cup \{0\} \tag{15}$$

L'objectif (12) consiste à trouver le rayon de couverture optimal.
Les contraintes (13) imposent que u_k vaut 0 si le nombre d'installations ouvertes dans le rayon D^k pour un client j est de 0. Sinon, si il y a au moins une installation dans ce rayon pour chaque client, alors u est libre.
Les contraintes (14) imposent qu'il y existe un unique k tel que u_k vaut 1. Ce u valant 1 sera pour l'indice k correspondant à la plus petite valeur possible de D, car l'objectif est en minimisation, et se fera parmi les u pouvant valoir 1. Ce sera donc pour le plus petit indice k vérifiant "il y a au moins une installation ouverte dans le rayon D^k pour chaque client", c'est-à-dire l'indice vérifiant D^k est le rayon de couverture optimal. Ainsi, u vaut bien 1 pour le rayon égal au rayon optimal, et 0 pour les autres.
Les contraintes (15) définissent les domaines des variables de décision.

3.2 Formulations prenant en compte les demandes et capacités (CpCP)

Maintenant, nous allons modifier les programmes précédents afin de prendre en compte les demandes et les capacités. L'objectif est de s'assurer que la demande totale des clients affectés à chaque installation ne dépasse pas sa capacité. On rappelle que la capacité de l'installation i est notée Q_i, et la demande d'un client j, q_j.

3.2.1 Version 1 : Formulation classique (CpCP$_1$)

Nous rajoutons une contrainte au modèle, afin de respecter les contraintes de capacité.

$$
\begin{aligned}
[P4] \quad & \min(1) \\
& \text{s.c. } (2), (3), (4), (5), (6), (7), (8) \\
& \sum_{j \in \mathcal{C}} q_j y_{ij} \leqslant Q_i x_i \qquad \forall i \in \mathcal{F} \qquad (16)
\end{aligned}
$$

Les contraintes (16) assurent que la demande totale des clients affectés à une installation ne dépasse pas la capacité de celle-ci, pour chaque installation. On multiplie la capacité d'une installation par x du côté droit de l'inéquation pour améliorer la relaxation linéaire. En effet, si dans la relaxation linéaire, x_i vaut 0.5, alors sa capacité sera diminuée en conséquence.

Il y a de la redondance avec les contraintes (4), car les contraintes (16) empêchent aussi un client d'être affecté à une installation non ouverte. On pourra tester la résolution du problème avec et sans les contraintes (4), afin de vérifier ou non l'utilité de garder les deux contraintes, dans la section 5.1.

3.2.2 Version 2 : Première formulation basée sur le rayon de couverture (CpCP$_2$)

On introduit dans le modèle la variable binaire y_{ij} pour chaque installation $i \in \mathcal{F}$ et pour chaque client $j \in \mathcal{C}$, qui vaut 1 si le client j est affecté à l'installation i, et 0 sinon.
On ajoute au modèle les contraintes imposant d'affecter chaque client à exactement une installation qui est ouverte, et les contraintes permettant de respecter les capacités des installations.
On ajoute aussi des contraintes assurant que chaque client soit affecté à une installation dans la limite du rayon de couverture.

$$
\begin{aligned}
[P5] \quad & \min(9) \\
& \text{s.c. } (2), (3), (4), (6), (7), (10), (11), (16) \\
& d_{ij} y_{ij} \leqslant D^0 + \sum_{k \in \mathcal{K}} (D^k - D^{k-1}) z_k \qquad \forall i \in \mathcal{F},\ \forall j \in \mathcal{C} \qquad (17)
\end{aligned}
$$

Les contraintes (17) assurent que chaque client soit affecté à une installation dont la distance n'est pas supérieure au rayon de couverture, celui-ci étant la somme du côté droit de l'inéquation, comme expliqué dans la section 3.1.2.

On remarque que cette formulation ressemble beaucoup à la première version, ce qui enlève tout l'intérêt du changement de variable. On peut écrire un autre modèle avec une autre formulation des contraintes assurant l'affectation des clients à des installations dans le rayon de couverture optimal, afin de conserver l'intérêt de cette version :

$$
\begin{aligned}
[P6] \quad & \min(9) \\
& \text{s.c. } (2),(3),(4),(6),(7),(10),(11),(16) \\
& \quad y_{ij} \leqslant \sum_{\substack{k\in\mathcal{K},\\ d_{ij}\leqslant D^k}} z_k \qquad \forall i\in\mathcal{F},\ \forall j\in\mathcal{C} \qquad (18)
\end{aligned}
$$

Avec les contraintes (18), si la distance entre l'installation i et le client j est inférieure ou égale au rayon optimal, alors la somme des z sera supérieure ou égale à 1, et donc y_{ij} sera libre. Sinon, si la distance est strictement supérieure au rayon optimal, alors la somme des z vaudra 0, et on a donc $y_{ij} \leqslant 0$. Ainsi, on ne peut donc affecter un client qu'à une installation dans la limite du rayon optimal.

On pourra comparer l'efficacité de ces deux formulations dans la section 5.2

3.2.3 Version 3 : Deuxième formulation basée sur le rayon de couverture (CpCP$_3$)

Comme pour le modèle précédent, on introduit dans le modèle les variables binaires y_{ij}, et on ajoute les contraintes imposant d'affecter chaque client à exactement une installation qui est ouverte, les contraintes permettant de respecter les capacités des installations, ainsi que les contraintes assurant que chaque client soit affecté à une installation dans la limite du rayon de couverture optimal.

$$
\begin{aligned}
[P7] \quad & \min(12) \\
& \text{s.c. } (2),(3),(4),(6),(7),(13),(14),(15),(16) \\
& \quad d_{ij}y_{ij} \leqslant \sum_{k\in\mathcal{K}\cup\{0\}} D^k u_k \qquad \forall i\in\mathcal{F},\ \forall j\in\mathcal{C} \qquad (19)
\end{aligned}
$$

Les contraintes (19) assurent que chaque client soit affecté à une installation dont la distance n'est pas supérieure au rayon de couverture, celui-ci étant la somme du côté droit de l'inéquation, comme expliqué dans la section 3.1.3.

Encore une fois, cette formulation est presque identique à la première version. Afin de conserver l'intérêt de cette formulation, on reformule les contraintes assurant l'affectation des clients à des installations dans la limite du rayon de couverture optimal.

$$
\begin{aligned}
[P7] \quad & \min(12) \\
& \text{s.c. } (2),(3),(4),(6),(7),(13),(14),(15),(16) \\
& \quad y_{ij} \leqslant 1 - \sum_{\substack{k\in\mathcal{K},\\ d_{ij}> D^k}} u_k \qquad \forall i\in\mathcal{F},\ \forall j\in\mathcal{C} \qquad (20)
\end{aligned}
$$

Avec les contraintes (20), si la distance entre l'installation i et le client j est inférieure ou égale au rayon optimal, alors la somme des u vaudra 0, et la valeur de y_{ij} sera libre. Sinon, la somme vaudra 1, et on aura donc $y_{ij} \leqslant 0$, c'est-à-dire qu'il n'est pas possible d'affecter le client j à l'installation i. Ainsi, on ne peut donc affecter un client qu'à une installation dans la limite du rayon optimal.

A nouveau, on va comparer l'efficacité de ces deux formulations dans la section 5.3.

4 Algorithmes de résolution

Pour écrire des algorithmes, le paquetage suivant est très utile : http://tug.ctan.org/macros/latex/contrib/algorithm2e/doc/algorithm2e.pdf

Idées d'heuristiques :
1. Avec les rayons de couverture. On part du plus petit, on regarde si c'est ok, et sinon, on augmente le rayon.
2. jsp

5 Expérimentations et résultats

La Table 1 présente un résumé des résultats de nos expérimentations.

TABLE 1 – Comparaison des différentes méthodes

Algorithme	Temps CPU (ms)		Gap MS		#Données MS
	Moyenne	Ecart type	Moyenne	Ecart type	
Heuristique PPV	29	96	8.4%	**36%**	6/38
Heuristique MI	16984	78093	**1.3%**	42%	**30/38**

5.1 Comparaison des deux formulations de CpCP$_1$

On compare la formulation de la section 3.2.1 et la même formulation, mais sans les contraintes (4).

5.2 Comparaison des deux formulations de CpCP$_2$

On compare les deux formulations de la section 3.2.2. La différence est l'écriture des contraintes empêchant d'affecter un client à une installation en dehors du rayon optimal.

5.3 Comparaison des deux formulations de CpCP$_3$

On compare les deux formulations de la section 3.2.3. La différence est l'écriture des contraintes empêchant d'affecter un client à une installation en dehors du rayon optimal.

6 Conclusion

C'est la fin.